

CLASSIFICATION OF WHEAT GROWING
ENVIRONMENTS IN OKLAHOMA

By

FELIX MARZA

Bachelor of Science
Technical University of Oruro
Oruro, Bolivia
1993

Master of Science
Altiplano National University
Puno, Peru
1998

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
August, 2001

CLASSIFICATION OF WHEAT GROWING
ENVIRONMENTS IN OKLAHOMA

Thesis Approved:

Burt F. Carver

Thesis Adviser

Eugene S. Krenyer, Jr.

Ernest C.

Arthur Klatt

Alfred H. Taylor

Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisor, Dr. Brett Carver, for his supervision and constructive guidance. I would like to thank and extend my appreciation to Dr. Gene Krenzer for the use of his data to conduct my research. My sincere appreciation extends to Dr. Arthur Klatt and Dr. Guihua Bai for their guidance, encouragement, and kindness.

Special thanks are given to the J. William Fulbright Foreign Scholarship Board and LASPAU (Academic and Professional Program for the Americas) for their generous financial support and assistance.

On a personal note, I would like to give my special thanks to my parents, Martin Marza and Rosa de Marza, for their strong encouragement at times of difficulty.

Finally, my gratitude is expressed to the Department of Plant and Soil Sciences for supporting me during these two years of study.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
II. LITERATURE REVIEW.....	3
III. MATERIALS AND METHODS.....	7
IV. RESULTS AND DISCUSSION.....	10
V. CONCLUSIONS.....	17
LITERATURE CITED.....	18
TABLES.....	20
FIGURES.....	24

LIST OF TABLES

Table	Page
1. Locations where the OSU wheat cultivar trials were conducted in Oklahoma from 1986 to 2000, including their average annual precipitation, number of years of testing, average grain yield, and yield standard deviation.....	20
2. Cultivars evaluated in OSU wheat cultivar trials from 1986 to 2000 including their source, number of years of testing, average grain yield, and yield standard deviation.....	21
3. Summary of factor loadings, variance explained by each factor, and final communality estimates for 21 environments in the OSU wheat cultivar trials.....	22
4. Summary of factor loadings, variance explained by each factor, and final communality estimates for 13 environments in the OSU wheat cultivar trials associated with area 1 (factor 1) in Table 3.....	23

LIST OF FIGURES

Figure	Page
1. Location of test sites for conducting the OSU wheat cultivar trials from 1986 to 2000.....	24
2. (a). Scatter plot of the first and second principal components representing scores for 21 environments. (b). Cultivar vectors with loadings for the first and second principal components.	25
3. Dendrogram for classification of 21 Oklahoma environments based on standardized wheat grain yield, complete linkage method of hierarchical agglomerative classification using squared Euclidean distance as the dissimilarity measure, and incremental sum of squares as the clustering strategy. (C = cluster).....	26
4. Trellis bar graph for mean standardized yields of 17 cultivars in five sets of environments delineated by cluster analysis.....	27

CHAPTER I

INTRODUCTION

Hard red winter wheat (*Triticum aestivum* L.) is the most important small grain grown in Oklahoma. It is extensively used for both forage for cattle and grain production from the same planting.

Replication of varieties and breeding lines over time and space is paramount to determining areas of adaptation. Representative locations are difficult to determine, especially if the target environment is variable. Genotype x environment interaction (GEI) complicates the identification of superior genotypes across a range of environments and represents changes in the relative performance of genotypes across different environments. To optimize wheat productivity, selection and identification of cultivars for appropriate production areas are needed. To reach this goal cultivars are assessed in multi-environment trials, and thus, the determination of appropriate locations becomes an important issue. If too few locations are chosen, the cultivars in the trial will not have been tested under the full range of conditions prevalent in the region. If too many locations are used, redundancy and waste of resources are at risk (Bradle and Arthur, 1992).

Several multivariate techniques can be used to address such issues. Principal component analysis (PCA) can be used to group environments into subsets where GEI within a group is minimized (Crossa et al., 1990). Cluster analysis, with classification and ordination of environments across years, can be accomplished by averaging squared

Euclidean distance values, which refers to the distance of dissimilarity in grain yield response between two locations. Relationships among test sites across years can also be estimated by factor analysis based on an average of all pair-wise correlations among test sites studied across years (Mirzawan et al., 1994). With factor analysis, a large number of correlated variables are reduced to a small number of main factors (Crossa et al., 1990).

This study was designed to understand genetic performance and environmental patterns unique to Oklahoma, which will allow us to identify key locations that uniquely discriminate among cultivars and advanced breeding lines. The analysis was conducted using 15 years of grain yield data from the Oklahoma State University - Oklahoma Cooperative Extension Service wheat cultivar trials.

LITERATURE REVIEW

Two or more genotypes planted in different environments may exhibit different relative performances or differences in scale among environments. This phenomenon is called “Genotype-Environment Interaction” (GEI) (Bradle and Arthur, 1992; Cornelius et al., 1993; Ouyang et al., 1995). Cooper et al. (1993) revealed that the primary objective in considering relationships among environments is to identify the degree of commonality among environments and patterns of discrimination among the genotypes, and as a consequence, identify a reduced testing regime.

Another study by Cooper et al. (1997) revealed that the types of target environments can be considered in relation to how environmental conditions impose stress upon genotypes. For example, some locations are more prone to water stress than others, or the soil nutritional status is inherently lower at some locations. Environments also have temporal elements where certain rainfall patterns occur with a degree of repeatability across years for a given location (Cooper and DeLacy, 1993).

Allard and Bradshaw (1964) found that interactions caused by weather variation in different years to be unpredictable. However, if the interactions are due to differences in soil types, and therefore associated with locations, they should be considered to be repeatable and predictable. Results suggested that cluster analysis grouped micro-environmental differences among locations, rather than macro-environmental differences (Bradle and Arthur, 1992; Yau et al., 1991).

There is substantial development in multivariate techniques to quantify and describe GEI in multi-environmental testing of genotypes. Genotype responses are multivariate rather than univariate, so multivariate techniques are in general more effective in explaining GEI than linear regression models (Oosterom et al., 1993). Variance components of GEI and phenotypic correlation of cultivar yields among test sites (as a measure of similarity of these sites) are two parametric examples. Powerful tools to explain GEI are principal component analysis, factor analysis, and cluster analysis based on cultivar differential yield response across environments (Peterson, 1992; Abdalla et al., 1996; DeLacy et al., 1994).

Principal component analysis (PCA) is a mathematical procedure that transforms a set of correlated response variables into a smaller set of uncorrelated variables called principal components (Johnson, 1998). In PCA, the environments can be conceptualized as a pattern in a G -dimensional space defined by the genotypes. The coordinates of each environment are determined from the yield of the G genotypes. The PCA defines a new set of coordinates in which a few orthogonal dimensions may account for most of the variance among environments (Crossa et al., 1993).

Factor analysis is a multivariate technique that explains the correlation structure among the measured variables (in this case, locations). In factor analysis, a large number of correlated variables are reduced to a small number of main factors. One basic objective of factor analysis is to determine whether the p response variables, or locations exhibit patterns of relationship with each other such that the variables can be partitioned into m subsets (locations with the same pattern), each consisting of a group of variables

tending to be more highly related to others within the subset than to those in other subsets (Johnson, 1998).

Cluster analysis is a technique used for combining observations into groups or clusters such that each group is homogeneous or compact with respect to certain characteristics. That is, observations in each group are similar to each other. Each group should be different from other groups with respect to the same characteristics (Sharma, 1996; Johnson, 1998).

Fox and Rosielle (1982) pointed out that observations may be environments or lines (genotypes), depending on whether the relationship among environments or among lines is being described. A genotype may be described in multidimensional space, with each dimension representing a test environment, the coordinates for which are the yields produced. Conversely, sites may be considered in multidimensional space with each dimension a genotype. In cluster analysis, the relative yields of a set of genotypes integrate the short-term interplay of biotic and abiotic influences during a crop cycle. Such agglomerative procedures, which fuse successively upwards from the level of the hierarchy until one group is formed, are not influenced by the number of fusions presented in the dendrogram. The truncation question concerns presentation or subsequent analysis of the groups formed and not clustering itself. Vertical ordering of groups in dendrograms is partly arbitrary, as the dendrogram can be considered a freely rotating mobile (with no specific order of locations). Although rotation changes the vertical order, it does not change the membership of a group (Romagosa et al., 1993).

Peterson and Pfeiffer (1989) used long-term performance data to allow more precise definition of site relationships, minimizing the effect of unusual or short-term

weather patterns or diseases. The nature of most long-term performance nurseries is such that the composition of cultivars is changing annually and test sites are not always represented each year. This makes the application of cluster analysis somewhat difficult. In this case factor analysis may provide a more effective means for understanding and describing location relationships. Johnson (1998) indicates that prior to performing any kind of multivariate analysis, principal component analysis (PCA) should be performed. PCA should be used mainly to screen the data, identify outliers, and to know the true dimension of the data.

The main objectives of this study were to determine groups of test sites in Oklahoma that represent similar environments and to identify the main factors that influence such groupings. Additionally, dual purpose and grain-only management systems were specifically compared to determine if these systems produce different yield patterns across locations and years. The analysis was based on grain yield responses of wheat cultivars grown in diverse conditions of moisture supply, temperature, soil type, and biotic/abiotic stresses in Oklahoma.

MATERIALS AND METHODS

The analysis in this study was based on grain yield of winter wheat cultivars measured in Oklahoma from 1986 to 2000, and reported by the Oklahoma Cooperative Extension Service. All experiments were arranged in the field in a randomized complete block design with four to six replicates. The number of cultivars evaluated each year varied from 17 to 23, with a total of 76 cultivars during the 15 years. The number of locations reporting data each year varied from 8 in 1989 to 22 in 2000, with a total of 41 locations over the 15 years. In this study we considered only those locations that were tested for three or more years as suggested by DeLacy et al. (1994) and Cooper et al. (1993). Therefore, we studied the relationship of 21 environments (Fig. 1). All environments were rainfed and managed for grain only, unless otherwise indicated as irrigated (I) or managed for dual-purpose (DP) of forage and grain. In order to avoid bias in overall yield response within locations in the same year, we considered only those cultivars that were tested in all locations per year.

Although yield data were reported in the same unit (kg/ha), sometimes it is easier to understand and compare when the response variables are standardized, thereby eliminating units of measurements (Steel et al., 1997). The transformation to Z scores was done through the following formula:

$$Z_{rj} = \frac{x_{rj} - \hat{\mu}_j}{\sqrt{\hat{\sigma}_{jj}}} \quad \text{for } r = 1, 2, \dots, N, \quad j = 1, 2, \dots, p$$

The variable Z_{rj} is called Z score for the j th location on the r th cultivar, x_{rj} is the yield

value for the j th location on the r th cultivar, $\hat{\mu}_j$ is the mean yield in the j th location, and $\sqrt{\hat{\sigma}_{jj}}$ is the variance (Cody and Smith, 1997). For the transformation we used the Standard procedure in the Statistical Analysis System (SAS) software.

The next step was to perform a principal components analysis. This procedure was applied to screen the data and identify or locate possible outliers in the data set. It served as a first attempt to group locations into subgroups of similar pattern. The PCA was performed on the correlation matrix from standardized data (Z scores). This step was done with the Princomp procedure of SAS.

Factor analysis was conducted to find relationships among environments. Phenotypic correlations for all pair-wise combinations of locations were determined for each year. Ordination of cultivars and environments was conducted on standardized grain yield data. Factor analysis was used to characterize similarities in cultivar responses among test sites from 15 years of yield data. The factor analysis was executed with the Factor procedure of SAS.

Cluster analysis was performed on standardized grain yield data, using the method of environmental classification for unbalanced data. With this method, the squared Euclidean distance (SED) values among environments obtained from each year were averaged over sets of data within and across years before the classification was done. The environments were classified using an agglomerative hierarchical classification procedure on the standardized data with squared Euclidean distance as a dissimilarity measure, and incremental sum of squares as a grouping strategy. Complete linkage was applied as a clustering method. The formula to calculate dissimilarity in standard Euclidean distance (ruler distance) between two observations was as follows:

$$D^2_{ij} = \sum (X_{ik} - X_{jk})^2 \text{ (Johnson, 1998),}$$

where D^2_{ij} is the squared distance between locations i and j , X_{ik} is the value of the k th cultivar for the i th location, and X_{jk} is the value of the k th cultivar for the j th location.

A relatively large distance between the last few clustering steps was an indicator of truncation of the clustering. We assumed that the cultivars tested in any given year were a representative sample of the adapted germplasm for winter wheat. Proximities among locations based on dissimilarities, measured by SED, was calculated for each year and averaged across years to produce a complete location x location proximity matrix. Matrices were averaged across years and weighted by the number of cultivars grown in each year, so that each year's contribution to the proximity measure was in proportion to the number of cultivars grown. Locations with missing cells in the weighted averaged proximity matrix were eliminated (Fox and Rosielle, 1982; Mirzawan et al., 1994; Basford et al., 1991; Abdalla et al., 1996). Cluster analysis was performed with the Cluster procedure of SAS.

CHAPTER IV

RESULTS AND DISCUSSION

Statewide average yields during 1986 to 2000 varied from 1630 kg/ha in 1995 to 3550 kg/ha in 1999. Individual environment yields varied from 270 kg/ha for Buffalo in 1995 to 6990 kg/ha for Goodwell-irrigated (I) in 1999. Average yields per location ranged from 1970 kg/ha for Chickasha-dual purpose (DP) to 4830 kg/ha for Goodwell-I (Table 1). Location standard deviations for yield ranged from 170 kg/ha for Forgan to 1030 kg/ha for Buffalo (Table 1). For the analysis period, Kingfisher was the only location tested every year. The average state yield for the 15-year period was 2800 kg/ha.

Cultivar average yields varied from 2550 kg/ha for Longhorn to 3330 kg/ha for 2174 (Table 2). Cultivar yield standard deviations ranged from 620 kg/ha for Tomahawk to 880 kg/ha for Custer. For the analysis period, Chisholm was the only cultivar tested every year.

Principal component analysis

Principal component analysis often precedes factor and cluster analysis to determine the relative importance of classification variables (Berdahl, et al, 1999). Eigenvalues from the first, second, third and fourth principal component axes, respectively, accounted for 42, 16, 13, and 6% of the total variance present. The first four vectors captured approximately 80% of the total variability. Thus, the 21-dimensional sample space of the complete data set can be reduced to a 4-dimensional

space of uncorrelated underlying variables, even though the new variables do not necessarily represent specific environments.

Further principal component analysis was performed with locations as experimental units and cultivars as variables (Fig. 2). The first two principal components (PC) explained a total of 92% of the total variation among cultivars. A bi-plot of PC1 (82%) and PC2 (10%) revealed five groups of locations (Fig. 2a). A circle was drawn around those locations with high similarity for loadings on principal components 1 and 2. The locations from right to left were grouped as follows: Group 1, Chickasha and Marshall; Group 2 with Haskell, Frederick, and Perkins; Group 3, comprised mainly of locations in north central Oklahoma, e.g., Kingfisher, Lahoma, Lamont but also Apache; Group 4, with Marshall-DP, Gage, and Tonkawa; and Group 5 with Forgan, Perkins-DP and Chickasha-DP. Locations Goodwell-I, Goodwell, Alva, Elk City-DP, Buffalo, and Cherokee-DP were plotted distantly from these defined groups. In Figure 2b, the cultivars are represented by vectors, of which their length indicate the proportion of the original variance explained by the first two principal components. The direction of the arrows indicates the relative loadings on the first and second principal components. The angle between two vectors indicates the correlation of standardized yield performance for two cultivars (lower angle indicates greater correlation). For example, one of the highest correlations was between cultivars 2174 and Dominator.

Factor analysis

Factor analysis divided the 21 locations in the OSU cultivar trials into four factors based on similarities in cultivar yield response (Table 3). These factors accounted for 77% of total variability in the correlation dependence structure among locations. Four

factors were considered optimal, because including an additional factor in the analysis accounted for less than 5.6% of the variability in the correlation matrix.

Factor loadings presented in Table 3 provide an approximate correlation between a location and a factor (a linear combination of the original variables or cultivars). Locations with a high loading on the same factor, or with similar loading patterns, are positively correlated and have relatively similar response patterns across environments. Twelve locations were associated with the first production area based on primary loadings on factor 1. This area corresponds to locations in the central and western part of the state, and contains the majority of the wheat growing area in Oklahoma. These locations appeared to react in a similar way to those factors responsible for variation in wheat yield over the 15-year period.

Locations Perkins-DP, Chickasha-DP, and Perkins showed high primary loadings on factor 2. Interestingly, the Perkins-DP and Chickasha-DP sites feature a forage-plus-grain production system, where forage is removed by clipping, not by grazing as for the other DP sites in other groups. Perkins also showed a high secondary loading for factor 3, and thus may serve as a transitional site among locations featuring grain-only management.

Locations with high primary loading for factor 3 were Marshall, Chickasha, and Frederick, which also had a high negative secondary loading for factor 2. Lamont, Alva, and Haskell were the locations with high primary loadings on factor 4, but all of them presented high secondary loadings. Lamont and Alva were transitional with locations in group 1, and Haskell was transitional with factor 2.

The large number of sites with secondary loadings suggests that production conditions change gradually throughout the state and true discrete production areas do not exist. Indeed, the factor analysis did not produce groups of locations with high geographic proximity, as demonstrated in group 3 (Marshall vs. Frederick) and group 4 (Alva vs. Haskell). Predominant secondary loadings among locations were with factor 4. Five locations showed secondary loadings > 0.30 with this factor.

The high number of locations with loadings on factor 1 warranted further subdivision into smaller production areas. All locations with primary factor 1 loadings or locations with secondary loadings > 0.40 for factor 1 were included in subsequent factor analysis (as recommended by Peterson, 1992). Therefore, we included Lamont whose secondary loading was 0.51 with respect to factor 1, but not Chickasha-DP and Alva.

Three smaller location groups were produced by this factor analysis (Table 4). Three factors accounted for 80% of the variability in the correlation dependence structure. There were eight locations with primary loadings on factor 1, where most of the locations were found transitional between factor 1 and 2 with the exception of Apache.

Forgan, Goodwell, and Goodwell-I were the only sites with primary loadings on factor 2, indicative of their geographic proximity. Forgan was not transitional. Elk City-DP and Kingfisher had high loadings on factor 3, but Kingfisher was transitional for factors 1 and 2.

Cluster analysis

Cluster analysis produced five groups of locations (2-5 locations per cluster), plus four non-classified locations that would be considered outliers (Fig. 3). The dendrogram

of the clustered locations was truncated at the five-cluster level. The truncation was based on Pseudo Hotelling's T^2 test, which produced the smallest value of 4.3 at the 5-cluster level.

Cluster 1 was dominated by locations in north central Oklahoma, with the exception of Apache, which unexpectedly clustered closely with Lahoma. Cluster 1 coincides with the major wheat cultivation area in the state, or approximately 30% of the statewide area for 2001 (Bloyd, 2001). Mean yields for cluster 1 as a whole were intermediate, exceeding those of cluster 4 and 5. Locations in cluster 1 are considered to have favorable conditions for wheat growth, with optimum rainfall (average of 750 mm) and silt and loam soil structure. One possible limiting factor in this area would be soil acidity, where the pH ranged from 5.4 for Lamont to 6.2 for Kingfisher (Zhang, 2000). OSU annual reports indicate that the disease pressure in this area mainly consisted of leaf rust, caused by *Puccinia triticina*.

Cluster 2 contained the locations Marshall and Chickasha (Fig. 3). These locations exhibit little eco-geographical relationship, but both have similar rainfall patterns (Table 1), where the tendency in the state is to decrease in annual amount from west to east.

Cluster 3 contained locations Perkins and Haskell from the eastern one-half of the state, and Elk City and Frederick from the southwest area. It is evident that these locations exhibited similar patterns for yield, but they also are subject to different stress patterns. Haskell and Perkins, because of their higher rainfall, are subject to disease problems, besides lower soil pH. On the other hand, based on rainfall patterns (Table 1), wheat yield in Elk City-DP and Frederick would be more limited by drought.

Cluster 4, with the lowest average yield, included three locations: Chickasha-DP, Perkins-DP, and Forgan. Certainly production conditions for the first two locations follow the same trend as mentioned above regarding forage removal by clipping. The low yield at Forgan, in comparison to other locations where the management system was grain-only, indicates that some underlying stress factor limited the realization of yield potential. One of the main factors would be drought conditions, as Forgan has relatively low annual precipitation (Table 1).

Cluster 5 contained locations Gage, Tonkawa, and Marshall-DP. The low yield potential of these locations would be attributed to different stresses. The low yield at Tonkawa might be due to low soil pH (5.3) or soil borne mosaic virus. The performance at Marshall-DP could be explained by the management system (dual purpose). At Gage, a high level of drought could be main source of stress explaining the relatively low yield.

Figure 4 compares the standardized cultivar grain yields within each of the five clusters found. This was accomplished through a Trellis bar plot to show similarities in cultivar responses among clusters. Cultivar performance in clusters 1 and 5 certainly showed high similarity in cultivar response, even though these clusters were distant from each other in the dendrogram (Fig. 3). According to Johnson (1998), this discrepancy may be explained by distortion of the dissimilarity measurement using Z scores, since standardization does not realistically illustrate the distance between clusters. Cultivars with outstanding performance in cluster 1 were 2174, Dominator, 2137, and Jagger; outstanding cultivars in cluster 5 were the same, with the exception of Jagger. Cultivar performance in cluster 2 was completely different from all other clusters. This tendency was found for factor analysis as well as for cluster analysis; thus, cultivar performance in

grain-only tests at Marshall and Chickasha was similar, but quite different from other sites. The cultivar with outstanding performance in this group was Cimarron, which, due to its high post-harvest seed dormancy, is not recommended for early-planted dual-purpose systems. Cultivar performance in cluster 3 was also different from other clusters. Outstanding cultivars for this group were 2137, Jagger, Custer, and Ike. Outstanding cultivars in cluster 4 were 2137, Chisholm, Cimarron, and Ike. Based on this analysis, they might be categorized as well adapted to forage removal, albeit by clipping.

CHAPTER V

CONCLUSIONS

Locations that consistently grouped together among the three multivariate methods (principal component analysis, factor analysis, and cluster analysis) were: 1) Apache, Lahoma, Lamont and Kingfisher; 2) Chickasha and Marshall (both grain-only); and 3) Perkins-DP and Chickasha-DP. Therefore we conclude that testing at every location within these groups would be redundant. Grouping of locations appeared to be associated with moisture supply (caused by natural weather patterns or by production system) rather than by geographic proximity.

Based on the three locations (Chickasha, Marshall, and Perkins) that featured both production systems (dual and grain purpose only), we found that cultivar responses to each management system changed among locations. As a result, grain-only or dual-purpose environments were not necessarily grouped together. On average, grain yields for dual-purpose production systems were reduced 35% from the grain-only production system.

The results indicated that long-term yield data can provide an effective way to determine relationships among production areas. The reporting of annual cultivar trials according to geographic zones does not appear entirely consistent with environmental response patterns revealed in this study by either of the three multivariate methods.

LITERATURE CITED

- Allard, R.W., A.D. Bradshaw. 1964. Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.* 4:503-508.
- Abdalla, O.S., J. Crossa, E. Autrique, and I.H. DeLacy. 1996. Relationships among international testing sites of spring wheat. *Crop Sci.* 36:33-40.
- Bloyd L.B. 2001. Oklahoma wheat varieties, crop year 2001. Statistical report. Oklahoma City, OK.
- Basford, K.E., P.M. Kroonenberg, and I.H. DeLacy. 1991. Three-way methods for multiattribute genotypeXenvironment data: an illustrated partial survey. *Field Crops Research* 27:131-157.
- Berdahl, J. D., H.F. Mayland, K.H. Assay, and P.G. Jefferson. 1999. Variation in agronomic and morphological traits among Russian wild rye accessions. *Crop Sci.* 39:1890-1895.
- Brandle, J.E. and L.M. Arthur. 1992. A comparative evaluation of two methods of selecting locations used for testing spring wheat cultivars. *Theor. Appl. Genet.* 83:301-304.
- Cody, R.P. and J.K. Smith. 1997. Applied statistics and the SAS programming language. 4th ed. Prentice-Hall, Inc. USA.
- Cooper, M., and I.H. DeLacy. 1993. Relationships among analytical methods used to study genotypic variation and genotype by environment interaction in plant breeding multi-environment experiments. *Theor. Appl. Genet.* 88:561-572.
- Cooper, M., D.E. Byth, and I.H. DeLacy. 1993. A procedure to assess the relative merit of classification strategies for grouping environments to assist selection in plant breeding regional evaluation trials. *Field Crops Research* 35:63-74.
- Cooper, M., R.E. Stucker, I.H. DeLacy, and B.D. Harch. 1997. Wheat breeding nurseries, target environments, and indirect selection for grain yield. *Crop Sci.* 37:1168-1176.
- Cornelius, P.L., D.A. Van Sanford, and M.S. Seyedsadr. 1993. Clustering cultivar into groups without rank-change interaction. *Crop Sci.* 33:1193-1200.
- Crossa, J., W.H. Pfeiffer, P.N. Fox, and S. Rajaram. 1990. Multivariate analysis for classifying sites, application to an international wheat yield trial. P. 214-233. In Manjit S. Kang (ed). *Genotype by environment interaction and plant breeding*. Baton Rouge, LA.

- Crossa, J., P.L. Cornelius, M. Seyedsadr, and P. Byrne. 1993. A shifted multiplicative model cluster analysis for grouping environments without genotypic rank change. *Theor. Appl. Genet.* 85:577-586.
- DeLacy, I.H., P.N. Fox, J.D. Corbett, J. Crossa, S. Rajaram, R. A. Fischer, and M. Ginkel. 1994. Long-term association of locations for testing spring bread wheat. *Euphytica* 72:95-106.
- Fox, P.N., and A.A. Rosielle. 1982. Reducing the influence of environmental main-effects on pattern analysis of plant breeding environments. *Euphytica* 31:645-656.
- Johnson, D.E. 1998. *Applied Multivariate Methods for Data Analysts*. Duxbury press. USA.
- Mirzawan, P.D.N., M. Cooper, and I.H. DeLacy. 1994. Retrospective analysis of the relationships among the test environments of the Southern Queensland sugarcane breeding programme. *Theor. Appl. Genet.* 88:707-706.
- Oosterom Van, E.J., D. Kleijn, S. Ceccarelli, and M.M. Nachit. 1993. Genotype by environment interaction of barley in the Mediterranean region. *Crop Sci.* 33:669-674.
- Ouyang, Z., R.P. Mowers, A. Jensen, S. Wang, and S. Zheng. 1995. Cluster analysis for genotype x environment interaction with unbalanced data. *Crop Sci.* 35:1300-1305.
- Peterson, C.J., and W.H. Pfeiffer. 1989. International winter wheat evaluation: relationships among test sites based on cultivar performance. *Crop Sci.* 29:276-289.
- Peterson, C. J. 1992. Similarities among test sites based on cultivar performance in the hard red winter wheat region. *Crop Sci.* 32:907-912.
- Romagosa, I. And P.N. Fox. 1993. Genotype x environment interaction and adaptation, In: *Plant breeding: Principles and prospects*. (eds M.D.Hayward, N.O. Bosemark, and I. Romagosa), Chapman & Hall. London, UK. pp 381-387.
- Sharma S. 1996. *Applied multivariate techniques*. Jonh Wiley & Sons Ed. USA. 493 p.
- Steel, R.G., J.H. Torrie, and D.A. Dickey. 1997. *Principles and procedures of statistics a biometrical approach*. 3rd ed. McGraw-Hill. USA.
- Yau, S.K., G. Ortiz-Ferrara, and J.P. Srivastava. 1991. Classification of diverse bread wheat-growing environments based on differential yield responses. *Crop Sci.* 31:571-576.
- Zhang, H. 2000. Oklahoma agricultural soil test summary 1994-1999. OSU, Current report, Oklahoma cooperative extension service. OK.

Table 1. Locations where the OSU wheat cultivar trials were conducted in Oklahoma from 1986 to 2000, including their average annual precipitation, number of years of testing, average grain yield, and yield standard deviation.

Locations	Average annual precipitation [†]	Years	Grain yield	
	mm		Mean	SD
		no.	-----kg ha ⁻¹ -----	
Alva	610	3	4130	240
Apache	762	14	2860	380
Buffalo	635	7	3110	1030
Cherokee-DP [†]	711	13	2750	540
Chickasha-DP	813	9	1970	300
Chickasha	813	3	3400	270
Elk City-DP	660	3	3220	210
Forgan	559	7	2110	170
Frederick	737	5	3080	180
Gage	559	4	2390	410
Goodwell-I [†]	432	4	4830	280
Goodwell	432	3	3530	530
Haskell	1067	12	2930	240
Kingfisher	787	15	2860	360
Lahoma	711	8	2910	230
Lamont	787	11	2690	260
Marshall-DP	762	8	2290	450
Marshall	762	3	3360	270
Perkins-DP	889	14	2050	250
Perkins	889	3	3000	260
Tonkawa	864	13	2420	310

[†] All locations were rainfed and managed for grain-only, unless indicated by I (irrigated) or by DP (managed as a dual purpose crop for forage plus grain)

Table 2. Cultivars evaluated in OSU wheat cultivar trials from 1986 to 2000 including their source, number of years of testing, average grain yield, and yield standard deviation.

Cultivars	Source	Years no.	Grain yield	
			Mean	SD
			-----kg ha ⁻¹ -----	
2137	KAES [†]	5	3240	710
2163	KAES	7	2890	760
2174	OAES [‡]	5	3330	740
7853	AGSECO	8	2940	680
Chisholm	OAES	15	2830	710
Cimarron	OAES	9	2810	790
Coronado	AgriPro	5	2930	810
Custer	OAES	6	3090	880
Dominator	Phillips Seed	3	3290	860
Jagger	KAES	6	3060	770
Karl 92	KAES	7	2770	730
Ike	KAES	6	2860	760
Longhorn	AgriPro	8	2550	740
Ogallala	AgriPro	6	2990	720
Oro Blanco	AgriPro	3	2930	830
Tomahawk	AgriPro	9	2820	620
Tonkawa	OAES	5	2780	750

[†] Kansas Agricultural Experimental Station

[‡] Oklahoma Agricultural Experimental Station

Table 3. Summary of factor loadings, variance explained by each factor, and final communality estimates for 21 environments in the OSU wheat cultivar trials.

Location	Primary factor loading	Secondary factor loading	Variance explained by each factor	Final communality estimates
<u>Group 1</u>	<u>Factor 1</u>		8.78 (42%)	
Tonkawa	0.95			0.96
Marshall-DP	0.91			0.84
Cherokee-DP	0.90	0.33-4†		0.97
Gage	0.86			0.77
Goodwell	0.83			0.72
Lahoma	0.81	0.41-4		0.84
Apache	0.77	0.31-4		0.70
Forgan	0.77	-0.51-4		0.90
Buffalo	0.66			0.50
Kingfisher	0.66			0.49
Goodwell-I	0.64			0.47
Elk City-DP	0.59	0.52-4		0.72
<u>Group 2</u>	<u>Factor 2</u>		3.33 (16%)	
Perkins-DP	0.90			0.89
Chickasha-DP	0.78	-0.35-1		0.73
Perkins	0.64	0.49-3		0.70
<u>Group 3</u>	<u>Factor 3</u>		2.66 (13%)	
Marshall	0.86			0.83
Chickasha	0.83			0.75
Frederick	0.66	-0.52-2		0.73
<u>Group 4</u>	<u>Factor 4</u>		1.35 (6%)	
Lamont	0.63	0.51-1		0.71
Alva	0.62	0.35-1		0.54
Haskell	0.51	0.30-2		0.35

† The number following the coefficient designates the secondary factor associations.

Table 4. Summary of factor loadings, variance explained by each factor, and final communality estimates for 13 environments in the OSU wheat cultivar trials associated with area 1 (factor 1) in Table 3.

Location	Primary factor loading	Secondary factor loading	Variance explained by each factor	Final communality estimates
<u>Group 1</u>	<u>Factor 1</u>		8.27 (64%)	
Apache	0.87			0.85
Cherokee-DP	0.83	0.36-3, 0.36-2†		0.94
Lahoma	0.78	0.40-3		0.83
Gage	0.76	0.48-2		0.81
Tonkawa	0.74	0.51-2, 0.30-3		0.90
Marshall-DP	0.67	0.51-2, 0.34-3		0.83
Lamont	0.59	0.44-3		0.55
Buffalo	0.49	0.32-2		0.43
<u>Group 2</u>	<u>Factor 2</u>		1.22 (9%)	
Forgan	0.99			1.00
Goodwell-I	0.48	0.35-1		0.43
Goodwell	0.56	0.51-3, 0.44-1		0.77
<u>Group 3</u>	<u>Factor 3</u>		0.97 (7%)	
Elk City-DP	0.95			0.98
Kingfisher	0.65	0.34-2, 0.31-1		0.63

† The number following the coefficient designates the secondary factor association.

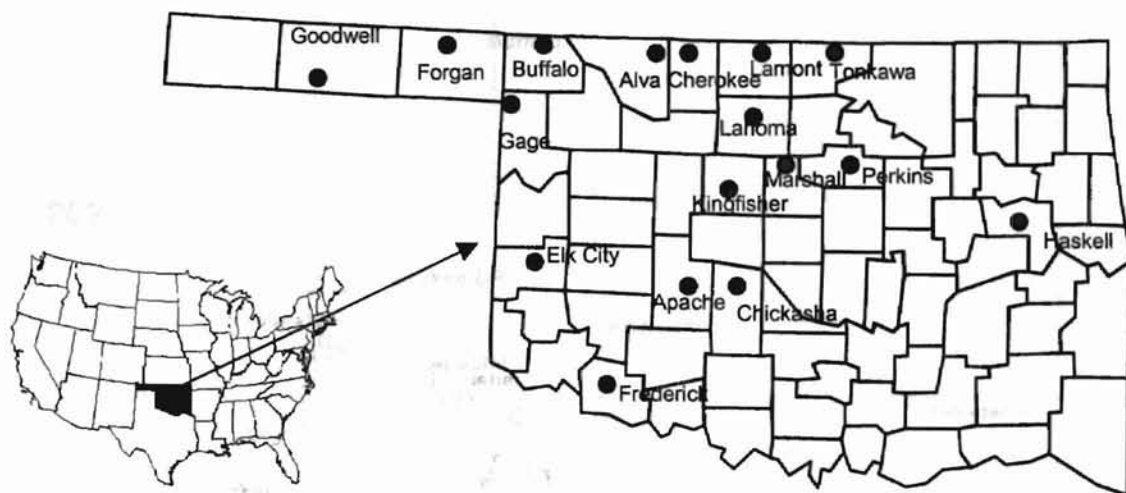


Fig. 1. Location of test sites for conducting the OSU wheat cultivar trials from 1986 to 2000.

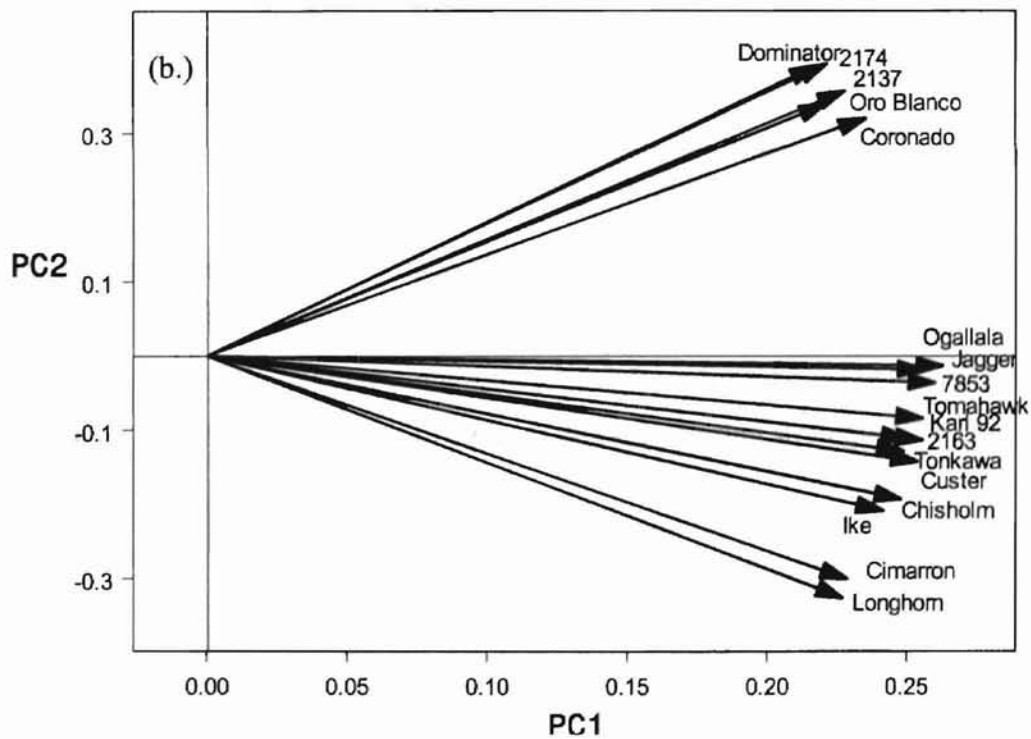
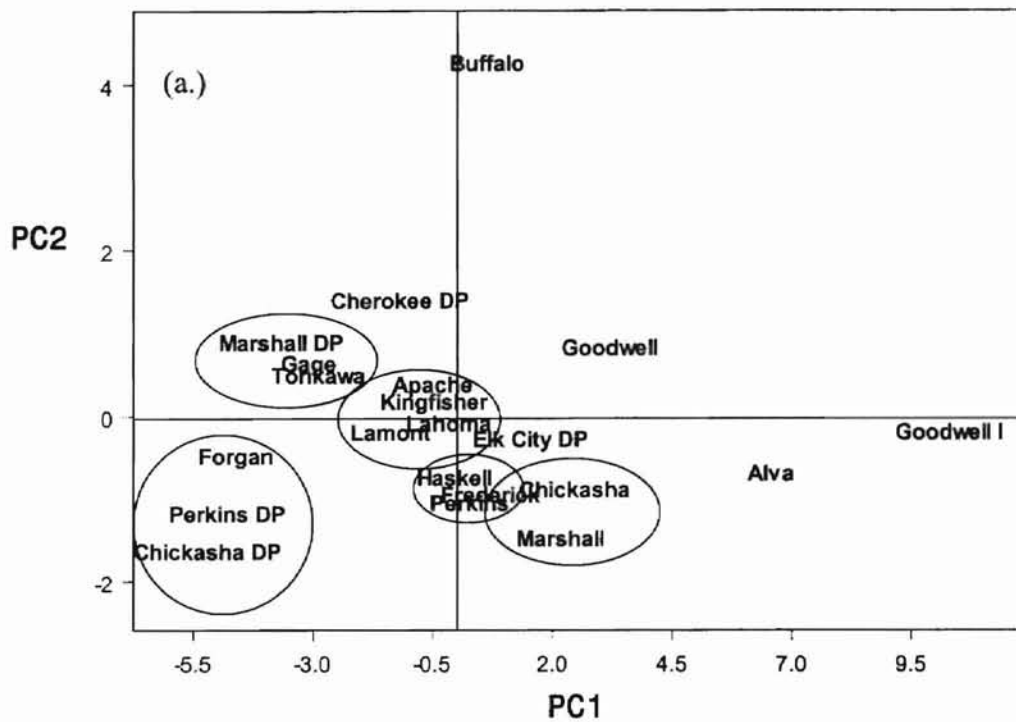


Fig. 2 (a). Scatter plot of the first and second principal components representing scores for 21 environments. (b). Cultivar vectors with loadings for the first and second principal components.

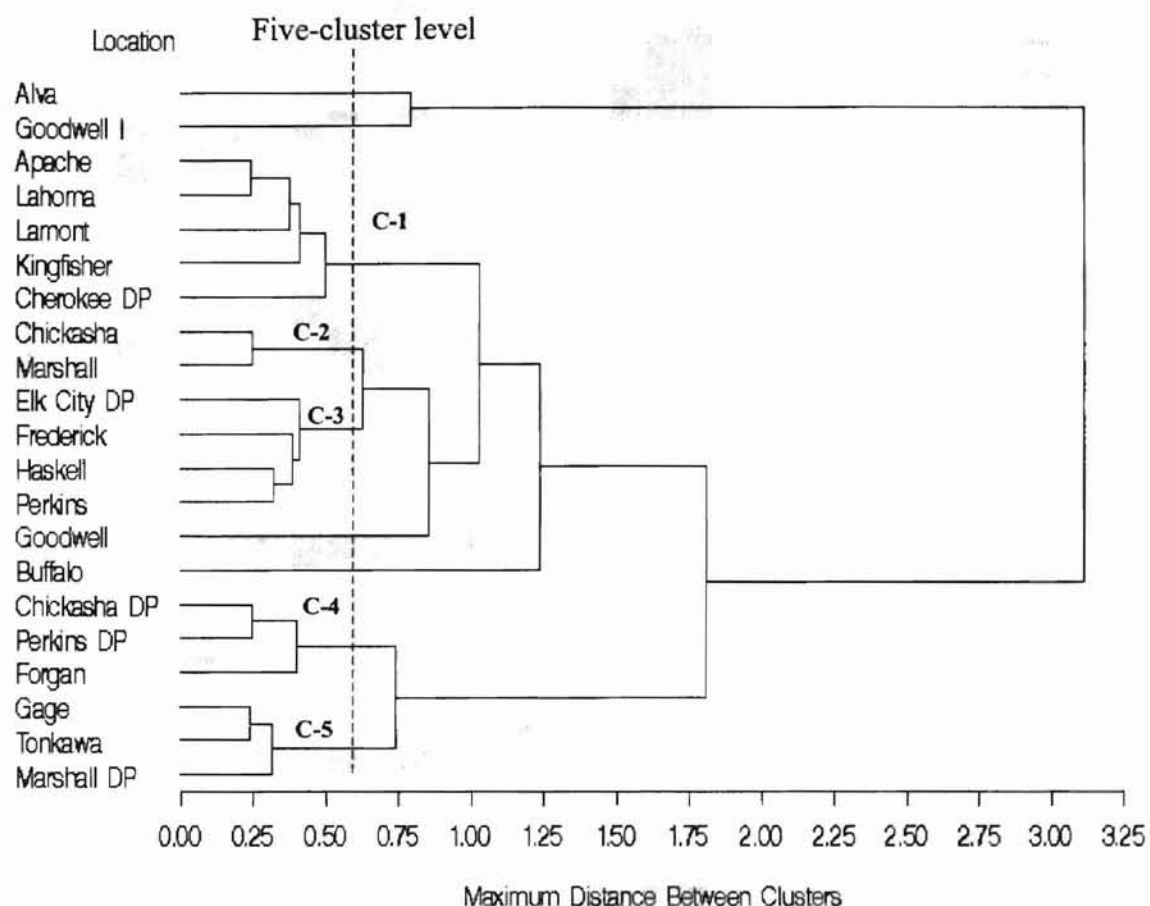


Fig. 3. Dendrogram for classification of 21 Oklahoma environments based on standardized wheat grain yield, complete linkage method of hierarchical agglomerative classification using square Euclidean distance as the dissimilarity measure, and incremental sum of squares as the clustering strategy. (C = cluster).

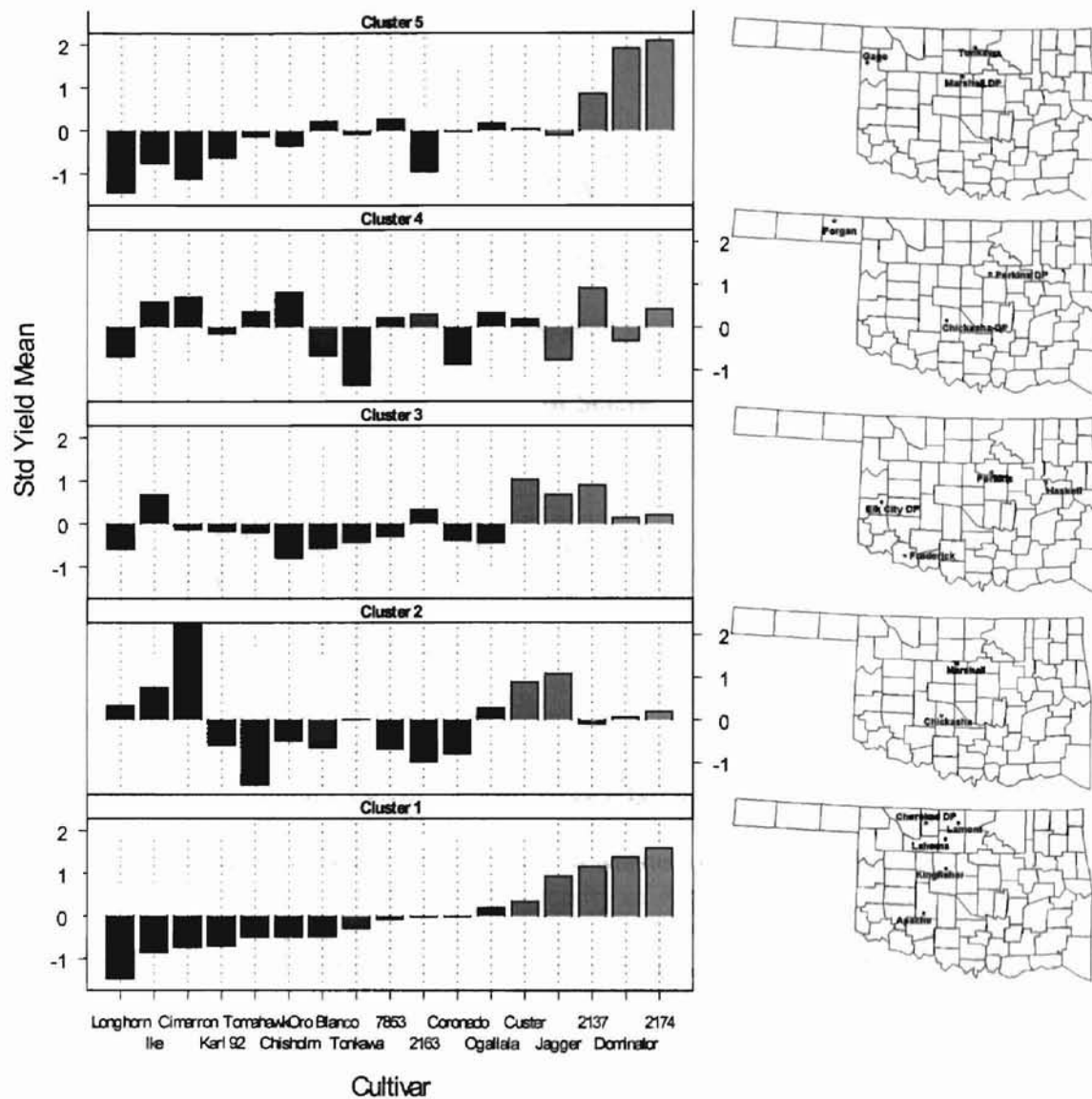


Fig. 4. Trellis bar graph for mean standardized yield of 17 cultivars in five sets of environments delineated by cluster analysis.

VITA 

Felix Marza

Candidate for the Degree of

Master of Science

Thesis: CLASSIFICATION OF WHEAT GROWING ENVIRONMENTS IN
OKLAHOMA

Major Field: Plant and Soil Sciences

Biographical:

Education: Graduated from Aniceto Arce High School, Oruro, Bolivia in 1984; received Bachelor of Science degree in Agronomy from Technical University of Oruro, Bolivia in 1993; received Master of Science degree in Andean Agriculture from National University of Altiplano, Puno, Peru in 1998. Completed the requirements for the Master of Science degree with a major in Plant and Soil Sciences at Oklahoma State University in August, 2001.

Experience: Employed by Technical University of Oruro, Department of Agronomy as a teacher and researcher in the field of Andean Genetic Resources, 1993 to present.